

PENGELOMPOKAN MINAT BELAJAR MAHASISWA MENGGUNAKAN TEKNIK *DATA MINING* DENGAN METODE *CLUSTERING*

Andri¹ dan Marlindawati²

Dosen Universitas Bina Darma

Jalan Jenderal Ahmad Yani No.03 Palembang

Sur-el: andri@mail.binadarma.ac.id¹, marlinda@gmail.com²

Abstract: *The development of information technology today allows data can be stored in digital form. At the college student population will continue to grow over time which causes the stored data, especially data related to students' academic will continue to grow. Data mining is an activity that aims to explore the information from a pile of data. Data mining is the process of data analysis using software to find patterns or certain rules of a number of large amounts of data that is expected to find knowledge to support keputusan. Dalam this research will be mining the data related to the academic students to be able to know the learning interest of students there is in college. This study will produce a model that can be used for grouping students based on their learning interests. Data mining techniques used in this study is the clustering technique.*

Keywords: *Information Technology, Data mining, and Clustering*

Abstrak: *Perkembangan teknologi informasi saat ini memungkinkan data dapat disimpan dalam bentuk digital. Pada Perguruan tinggi jumlah mahasiswa akan terus bertambah seiring berjalannya waktu yang menyebabkan data yang tersimpan terutama data yang berhubungan dengan akademik mahasiswa akan terus bertambah.. Data mining merupakan kegiatan yang bertujuan untuk menggali informasi dari tumpukan data yang besar. Data mining merupakan proses analisis data menggunakan perangkat lunak untuk menemukan pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat menemukan pengetahuan guna mendukung keputusan. Dalam penelitian ini akan dilakukan mining data yang berhubungan dengan akademik mahasiswa untuk dapat mengetahui minat belajar dari mahasiswa yang ada dalam perguruan tinggi. Penelitian ini akan menghasilkan sebuah model yang dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajarnya. Teknik Data mining yang digunakan dalam penelitian ini adalah teknik clustering.*

Kata Kunci: *Teknologi Informasi, Data mining, dan Clustering*

1. PENDAHULUAN

Perkembangan teknologi informasi terutama yang berhubungan dengan media penyimpanan saat ini berkembang begitu pesat. Media penyimpanan data dapat digunakan oleh sebuah perguruan tinggi untuk menyimpan data-data yang berhubungan dengan akademik mahasiswa. Data bagi suatu perguruan tinggi sangat bermanfaat jika diolah lebih lanjut untuk menghasilkan sebuah informasi yang dapat digunakan dalam pengambilan keputusan bagi pucuk pimpinan perguruan tinggi tersebut.

Data merupakan sumber utama yang dapat dimanfaatkan untuk menghasilkan informasi. Informasi begitu penting bagi sebuah perguruan tinggi untuk dapat tetap bertahan dalam menghasilkan lulusan yang terbaik. Pengolahan data yang maksimal maka akan menghasilkan informasi yang berkualitas.

Seiring dengan berjalannya waktu, data dalam sebuah perguruan tinggi akan terus bertambah sehingga akan menjadi tumpukan data yang menggunung. Jika tidak dikelola dengan baik maka data tersebut hanya berupa data yang tidak ada manfaatnya. Koleksi data yang mempunyai sifat berorientasi subjek,

terintegrasi, time-variant, dan bersifat tetap dari kumpulan data dalam mendukung proses pengambilan keputusan manajemen disebut dengan *Data Warehouse* (Inmon & Richard D.H, 1994). Sedangkan menurut Vedette Poe (1997), *data warehouse* merupakan *database* yang bersifat analisis dan *read only* yang digunakan sebagai fondasi dari sistem penunjang keputusan. Sumber lain menyebutkan *data warehouse* merupakan *database* relasional yang didesain lebih kepada *query* dan analisis daripada proses transaksi, biasanya mengandung *history* data dari proses transaksi dan bisa juga data dari sumber lainnya (Lane, 2002). Menurut Ponniah (2001) *data warehouse* adalah suatu paradigma baru dilingkungan pengambilan keputusan strategik. *Data warehouse* bukan suatu produk tetapi suatu lingkungan dimana *user* dapat menemukan informasi strategik. *Data warehouse* adalah kumpulan data-data logik yang terpisah dengan *database* operasional dan merupakan suatu ringkasan. Dengan memanfaatkan teknik *Data mining* maka akan didapatkan informasi yang berguna dari tumpukan data yang ada yang sudah menjadi *data warehouse* tadi.

Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies & Paul, 2004). Menurut Pramudiono (2007) *Data mining* dapat juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari kumpulan data.

Data mining dapat dimanfaatkan untuk menggali informasi dari tumpukan data

akademik yang berhubungan dengan mahasiswa untuk menemukan informasi atau pola dari minat belajar mahasiswa. Informasi yang didapat dari teknik mining data ini adalah sebuah keputusan yang dapat dimanfaatkan pihak yang berwenang dalam perguruan tinggi untuk mengarahkan konsentrasi mahasiswa tersebut.

Berdasarkan fenomena inilah maka akan dilakukan sebuah penelitian yang bertujuan untuk membuat sebuah model *Data mining* yang dapat digunakan untuk melakukan pengelompokan mahasiswa berdasarkan minat belajarnya. Data yang digunakan dalam penelitian ini adalah data Mahasiswa Program Studi Sistem Informas Fakultas Ilmu Komputer Universitas Bina darma tahun angkatan 2011-2013 dan Teknik *Data mining* yang digunakan yaitu *Clustering*.

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah pengamatan (*observasi*), wawancara (*interview*) dan studi pustaka.

2.2 Karakteristik Data Warehouse

Karakteristik *data warehouse* terdiri dari:

- 1) Berorientasi subyek

Data warehouse adalah tempat penyimpanan berdasarkan subyek bukan

berdasarkan aplikasi. Subyek merupakan bagian dari suatu perusahaan.

2) Data yang terintegrasi

Sumber data yang ada dalam *data warehouse* tidak hanya berasal dari *database* operasional (*internal source*) tetapi juga berasal dari data diluar sistem (*external source*).

3) *Nonvolatile*

Data dalam *database* operasional akan secara berkala atau periodik dipindahkan kedalam *data warehouse* sesuai dengan jadwal yang sudah ditentukan. Sekali masuk ke dalam *data warehouse*, data adalah *read-only*.

4) *Time-Variant*

Sistem operasional mengandung data yang bernilai sekarang sedangkan data dalam *data warehouse* mengandung data tidak hanya data terkini tetapi juga data *history* yang akan digunakan dalam analisis dan pengambilan keputusan.

5) Ringkas

Jika diperlukan, data operasional dikumpulkan ke dalam ringkasan-ringkasan.

6) *Granularity*

Pada data warehouse menganalisis harus memperhatikan *level-of-detail* misalkan perhari, ringkasan perbulan, ringkasan per-tiga-bulan.

7) Tidak ternormalisasi

Data di dalam sebuah *data warehouse* biasanya tidak ternormalisasi dan sangat redundan.

Data warehouse mengandung beberapa elemen penting antara lain (Mallach, 2000):

1) Sumber data yang digunakan oleh *data warehouse*, database transaksional dan sumber data eksternal.

2) Proses ETL (*Extraction, Transformation, Loading*) dari sumber data ke *database data warehouse*.

3) Membuat suatu ringkasan atau *summary* terhadap *data warehouse* misalkan dengan menggunakan fungsi agregat.

4) Metadata, metadata mengacu data tentang data. Metadata menguraikan struktur dan beberapa arti tentang data, dengan demikian mendukung penggunaan efektif atau tidak efektif dari data.

5) *Database data warehouse*.

6) *Database* ini berisi data yang detail dan ringkasan data dari data yang ada di dalam *data warehouse*.

7) *Query Tools* yaitu dengan OLAP (*Online Analytical Processing*) dan *data mining*. *Tool* untuk *query* ini meliputi antarmuka pengguna akhir dalam mengajukan pertanyaan kepada *database*, dimana proses ini disebut sebagai *On-line Analytical Processing* (OLAP).

8) *User*, merupakan pengguna yang memanfaatkan *data warehouse* tersebut.

9) Proses ETL(*Extraction, Transformation, dan Loading*) merupakan proses yang harus dilalui dalam pembentukan *data warehouse*, yaitu:

a. *Extraction*

Ekstraksi dilakukan dari sumber data yang digunakan melalui proses

pemilihan data yang kemudian disimpan pada basis data sementara. Dalam penelitian ini, basis data sementara ditempatkan pada penyimpanan basis data, mesin dan *platform* yang sama dengan yang digunakan untuk *data warehouse* nantinya.

b. *Cleaning*

Proses *cleaning* adalah proses untuk membersihkan data yang rangkap, tidak konsisten, rusak, tidak benar (bernilai salah), data yang hilang (kosong), atau data dengan spasi yang berlebihan.

c. Transformasi

Proses transformasi yang dilakukan dibagi berdasarkan dua *level* yaitu *record level* dan *field level*. Pada proses ini dilakukan proses pemilihan, penggabungan dan agregasi untuk mendapatkan data ringkasan sesuai dengan dimensi yang akan dibuat.

d. *Loading*

Proses terakhir yang perlu dilakukan adalah proses pemuatan data (*loading*). Data yang digunakan pada tahap ini merupakan data dari proses-proses yang dilakukan sebelumnya yaitu ekstraksi, pembersihan (*cleaning*) dan transformasi untuk kemudian dimasukkan ke dalam *data warehouse*. Cara pemuatan data ke dalam *data warehouse* adalah dengan menggunakan *script* yang dijalankan secara periodik.

2.3 *Data Mining*

Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Davies, 2004). *Data mining* dapat juga disebut proses menemukan suatu pola dalam data. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan dan data yang dibutuhkan dalam jumlah besar (Witten, 2005).

Menurut Davies (2004), *Data mining* memiliki beberapa karakteristik, yaitu:

- 1) *Data mining* berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- 2) *Data mining* biasanya menggunakan data yang sangat besar yang digunakan untuk membuat hasil yang lebih dipercaya.
- 3) *Data mining* berguna untuk membuat keputusan yang kritis, terutama dalam strategi

Data mining juga disebut dengan *Knowledge Discovery in Database (KDD)*, yaitu kegiatan yang meliputi pengumpulan, pemakaian data untuk menemukan keteraturan, pola dan hubungan dalam set data yang berukuran besar. Keluaran *data mining* dapat dipakai untuk pengambilan keputusan. *Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. (Pramudiono, 2007). *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan

pengambilan informasi dari *database* yang besar. (Larose, 2005).

Secara garis besar *datamining* dapat dikelompokkan menjadi 2 kategori utama, yaitu (Tan *et al*, 2006):

- 1) *Descriptivemining*, yaitu proses untuk menemukan karakteristik penting dari data dalam suatu basis data. Teknik *Data mining* yang termasuk dalam *descriptive mining* adalah *clustering*, *association*, dan *sequential mining*.
- 2) *Predictive*, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variabel di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi

2.4 Tahapan dalam *Data Mining*

Menurut Han (2006) terdapat enam tahap dalam *data mining*:

- 1) Pembersihan data (*cleaning*)
- 2) Integrasi data (*integration*)
- 3) Seleksi data (*selection*)
- 4) Transformasi data (*transformation*)
- 5) Proses mining (*Data mining*)
- 6) Evaluasi pola (*pattern evaluation*)
- 7) Presentasi pengetahuan (*knowledge presentation*)

2.5 *Business Intelligence Development Studio (BIDS)*

Business Intelligence Development Studio (BIDS) adalah *tool* (alat bantu) yang digunakan untuk mendesain, membuat, dan

mengetes *package*. BIDS menyediakan sebuah lingkungan pengembangan yang berorientasi grafis yang membantu proses pengkopian, pengaturan dan pembuatan *package* menggunakan menu dan kotak *Tool* dengan metode *drag-and-drop*. beberapa *tools* yang ada dalam BIDS yang digunakan dalam penelitian meliputi:

1) *SQL Server Integration Service (SSIS)*

SQL Server Integration Service merupakan sebuah *tools* yang digunakan untuk melakukan proses ETL (*Extract, Transform, and Load*).

2) *SQL Server Analysis Service (SSAS)*

SSAS merupakan *tools* yang digunakan untuk melakukan proses OLAP dan *Data mining*.

2.6 *Clustering*

Clustering atau analisis kelompok merupakan pekerjaan mengelompokkan data (objek) berdasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya (Tan, 2006). Tujuan dari pengelompokan ini adalah agar objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok yang lain.

Tujuan pekerjaan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan (Prasetyo, 2012). Jika tujuannya untuk pemahaman, kelompok yang terbentuk harus menangkap

struktur alami data, biasanya proses pengelompokan dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau summarization (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok yang untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

2.7 Metode Analisis *K-Means*

Dalam statistic dan mesin pembelajaran, pengelompokan *K-Means* merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan ke dalam K kelompok (*cluster*) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan *mean* (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* dimana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma. Pengelompokan data dengan metode *K-Means* ini secara umum dilakukan dengan langkah berikut (Prasetyo, 2012):

- 1) Tentukan jumlah kelompok
- 2) Alokasikan data ke dalam kelompok secara acak
- 3) Hitung pusat kelompok (sentroid/rata-rata)

dari data yang ada di masing-masing kelompok

- 4) Alokasikan masing-masing data ke sentroid/rata-rata tersebut
- 5) Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok, atau apabila ada perubahan nilai sentroid diatas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

3. HASIL DAN PEMBAHASAN

Untuk melakukan mining terhadap data maka dibutuhkan beberapa tabel. Tabel ini dalam database sebenarnya sudah ada, namun tabel yang dibuat ini lebih kecil, karena hanya terdiri dari field-field yang diperlukan untuk proses *Data mining* saja. Rancangan terhadap database yang akan digunakan untuk proses dijelaskan di bawah ini.

3.1 Rancangan Database

Database yang dirancang dalam penelitian ini terdiri dari beberapa tabel. Tabel-tabel dalam rancangan dapat diuraikan sebagai berikut:

- 1) Tabel Mahasiswa (tb_mhs)

Tabel ini digunakan untuk menampung data-data pribadi mahasiswa. Adapun atribut (*field*) dapat diuraikan sebagai berikut:

Tabel 1. Tabel Mahasiswa

Nama	Type	Width
nim*	varchar	9
nama	varchar	40
jenis_kelamin	varchar	2
temp_lahir	date	50
tgl_lahir	date	10
kd_progdi	varchar	3
alamat	varchar	250
asal_sek	varchar	20
kota	varchar	25
pek_ortu	varchar	20

2) Tabel KHS (tb_khs)

Tabel 2. Tabel Kartu Hasil Studi (KHS)

Nama	Type	Width
Nim*	varchar	
kd_mk	varchar	9
sms	varchar	10
sms_pendek	varchar	1
tahun_akademik	varchar	1
kelas	varchar	9
kd_dosen	varchar	5
kd_progdi	varchar	5
tugas	int	3
kuis	int	6
mid	int	6
semester	int	6
nilai_angka	int	6
nilai_huruf	varchar	1

3) Tabel Matakuliah (tb_mk)

Tabel 3. Tabel Matakuliah

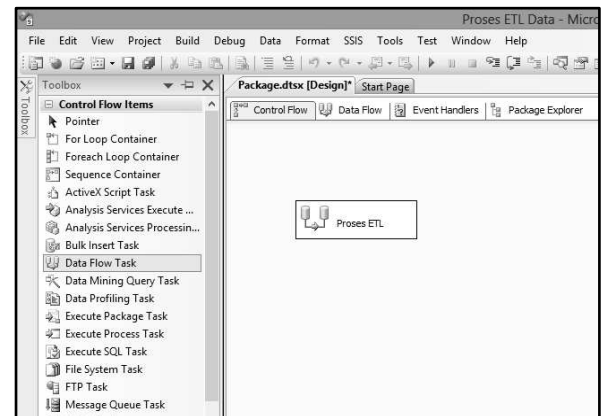
Nama	Type	Width
kd_mk*	varchar	10
nm_mk	varchar	30
sms	varchar	1
sks	int	1
kd_progdi	varchar	3

3.2 ETL (Extract, Transform and Load)

ETL merupakan proses yang bertujuan untuk melakukan pembentukan data warehouse yang bersumber dari data mentah dalam format

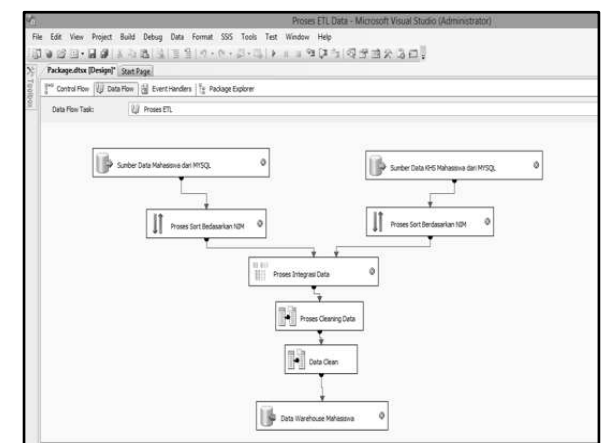
database MySQL. Data yang di ekstrak akan dilakukan proses transform dan load data kedalam format baru yaitu format mdb. Proses ETL ini dilakukan dengan menggunakan *tools SQL Server Integration Service (SSIS)* yang ada dalam *Business Intelligence Development Studio (BIDS)*. ETL yang dilakukan dalam penelitian dapat dilihat pada gambar berikut ini:

1) Pembuatan Control Objek



Gambar 1. Control Objek

2) Proses ETL



Gambar 2. Proses ETL

3.3 Seleksi Data (Data Selection)

Sumber data yang digunakan dalam penelitian berasal dari datamahasiswa tahun angkatan 2011 sampai dengan 2013 pada jurusan

sistem informasi. Beberapa jenis data diperoleh dari sistem yang berjalan namun hanya data mahasiswa dan data IPK saja yang digunakan untuk penelitian, dikarenakan informasi yang terkandung di dalamnya sudah mewakili informasi yang dibutuhkan untuk dijadikan indikator penentu dalam pengelompokan data keluaran yang diinginkan.

Data yang diperoleh adalah sebanyak 1.000 *record* data yang berasal dari *tb_mhs*, 4.1367 *record* data yang berasal dari *tb_khs* dan 417 *record* berasal dari *tb_mk*. *Dataset* mahasiswa terdiri dari 23 atribut yang menjelaskan identitas diri mahasiswa dan informasi tentang keadaan mahasiswa yang bersangkutan. Atribut data mahasiswa diambil dari penggabungan beberapa tabel yang terdiri dari *tb_mhs*, *tb_khs*, dan *tb_mk*. Berikut ini adalah tahap-tahap yang akan digunakan dalam penelitian ini.

1) Data Mahasiswa (tb_mhs)

2) Data KHS (tb_khs)

Gambar 4. Record Tabel Kartu Hasil Studi

3) Data Matakuliah (tb_mk)

Gambar 5. Tabel Matakuliah

3.4 Praproses Data

3.4.1 Pembersihan Data (Data Cleaning)

Tahap kedua pada proses *Data mining* adalah *cleaning data* yaitu melakukan pembersihan data terhadap *noise* yang ditemukan berupa *missing value*, *inkonsisten data*, dan *redundant data*.

mining menggunakan teknik *clustering* dan algoritma K-Means.

4. SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dijelaskan dapat disimpulkan bahwa perancangan dan Implementasi Aplikasi *data mining* untuk mengelompokkan minat belajar mahasiswa dalam menentukan konsentrasi matakuliah pilihan program studi sistem informasi Fakultas Ilmu Komputer Universitas Bina Darma Palembang dapat membantu mahasiswa dalam menyalurkan minat mereka sehingga sesuai dengan kemampuan dan bakat yang dimiliki.

DAFTAR RUJUKAN

- Davies and Paul Beynon. 2004. *Database Systems Third Edition*. Palgrave Macmillan. New York.
- Han, J., & Kamber, M. 2006. *Data Mining Concepts and Techniques, Second Edition*. Morgan Kauffman. San Fransisco.
- Inmon, William H., Richard D. Hackathorn. 1994. *Using the Data Warehouse*. John Wiley & Son's.
- Lane, Paul. 2002. *Oracle9i Data Warehousing Guide*. Oracle Corporation.
- Larose D, T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. Jhon Wiley & Sons Inc.
- Mallach, Efrem G. 2000. *Decision Support and Data Warehouse Systems*. Irwin McGraw Hill. Singapore.
- Ponniah, Paulraj. 2001. *Data Warehouse Fundamentals: a Comprehensive Guide for IT Professional*. John Wiley & Sons. New York.
- Pramudiono, I. 2007. *Algoritma Apriori*. (Online). (Diakses <http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip&1172210143>, 12 Mei 2015).
- Prasetyo, Eko. 2012. *Data Mining KONSEP dan Aplikasi menggunakan MatLab*. Andi. Yogyakarta.
- Tan, P., Steinbach, M., Andkumar,V. 2006. *Introduction to Data mining*. Pearson Addison Wesley Boston.
- Vedette Poe, 1997. *Building A Data Warehouse for Decision Support (2nd Edition)*. Prentice Hall.
- Witten, I. H and Frank, E. 2005. *Data mining : Practical Machine Learning Tools and Techniques Second Edition*. Morgan Kauffman. San Francisco.